Perspectives in Drug Discovery and Design, **17:** 97–118, 1999.
KLUWER/ESCOM

Solution X-ray scattering as a probe of hydration-dependent structuring of aqueous solutions

GREG HURA[a], JON M. SORENSON[b], ROBERT M. GLAESER[c] and TERESA HEAD-GORDON[d,*]

[a] *Graduate Group in Biophysics, University of California at Berkeley and Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.*
[b] *Department of Chemistry, University of California at Berkeley, Berkeley, CA 94720, U.S.A.*
[c] *Department of Molecular and Cell Biology, University of California at Berkeley, and Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.*
[d] *Physical Biosciences and Life Sciences Divisions, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.*

Received 20 February 1999; Accepted 4 May 1999

Summary. We report on new X-ray solution scattering experiments and molecular dynamics simulations conducted for increasing solute concentrations of *N*-acetyl-amino acid-amides and -methylamides in water, for the amino acids leucine, glutamine, and glycine. As the concentration increases, the main diffraction peak of pure water at Q = 2.0 $Å^{-1}$ shifts to smaller angle for the larger leucine and glutamine amino acids, and a new diffraction peak grows in at Q $\sim$ 0.8 $Å^{-1}$ for only the hydrophobic amino acid leucine. The unaltered value of the peak position at Q $\sim$ 0.8 $Å^{-1}$ over a large concentration range suggests that a stable and ordered leucine solute–solute distribution is sustained. Simulations of the distributions of leucines in water that reproduce the experimental observable show that mono-dispersed to small molecular aggregates of two to six hydrophobic amino acids are formed, as opposed to complete segregation of the hydrophobic solutes into one large cluster. The scattering results for the hydrophobic leucine amino acid are contrasted with experiments and simulations of the model hydrophilic side chain glutamine and the model backbone glycine. The self-assembly process of protein folding modeled with these experiments, in particular the condensation to a hydrophobic core, shares similar issues with the desolvation phenomena that are important in drug discovery.

Key words: desolvation, energy functions, hydrophobicity, protein folding, solution scattering

Introduction

The influence of aqueous solvent on the molecular recognition of drug substrates will be determined by the hydrophobic and hydrophilic character of the receptor sites and substrate, positions of the water binding sites, steric

* To whom correspondence should be addressed. E-mail: TLHead-Gordon@lbl.gov

effects, and the thermodynamics of 'desolvating' the ligand and protein active site in order to 'dock' the drug [1]. Crystallography has provided structural detail about protein active sites, complexes that they form with substrates or inhibitors, as well as sometimes characterizing the water binding geometries [2–4]. Transfer free energies of model compounds from organic liquids into aqueous solvent, and/or electrostatic effects determined from a solution to the Poisson–Boltzmann equation, have been used to provide for some estimate of solvation free energy changes on docking [5–8]. To the extent that these approaches are lacking – either because of accuracy or computational cost – the discourse of this volume is to explore alternative approaches that might give new insight and ultimately promote improved solvation models that better predict drug binding affinity.

In a series of papers we have combined molecular dynamics simulations [9–13], neutron and X-ray solution scattering experiments [10–14], and protein folding models [15] to deepen our understanding of hydration by placing greater emphasis on the varied chemical properties of the individual amino acids and on the complexity of the hydration interaction. Our work can be viewed as a model systems approach of determining the solution structure and free energy of amino acid association, beginning with early protein folding events when the concentration of amino acid solutes is dilute, to later stages of collapse when the increasing concentration contributes to the formation of a hydrophobic core [13].

The self-assembly process of protein folding, in particular the condensation to a hydrophobic core, shares similar issues in desolvation phenomena that are important in drug discovery. With our approach, an appropriately designed combination of solution scattering experiments and simulations on amino acids can also be used to characterize hydration structure or free energy of stabilization of hydrophobic or hydrophilic solute distributions. Information about hydration for these systems can be useful in understanding the solvation state of the protein active site, or in relation to similar functional groups of the amino acid solutes that are shared with the drug of interest.

In this article, we describe X-ray solution scattering results on the behavior of *N*-acetyl-leucine-methylamide (NALMA) and *N*-acetyl-leucine-amide (NALA) in water as the concentration of the blocked amino acid increases. The results for this hydrophobic amino acid are contrasted with the results obtained for a hydrophilic amino acid, *N*-acetyl-glutamine-amide (NAQA), and the model backbone *N*-acetyl-glycine-methylamide (NAGMA). As the concentration increases, the main diffraction peak of pure water at Q = 2.0 $Å^{-1}$ (where Q is the momentum transfer, related to the scattering angle as $Q = 4\pi \sin(\theta/2)/\lambda$) shifts to smaller angle for NALMA, NALA, and NAQA. In addition, a new diffraction peak grows in at $Q \sim 0.8$ $Å^{-1}$ for only the hydrophobic

amino acids NALMA and NALA. The peak position at Q $\sim$ 0.8 Å^{-1} remains stationary for a concentration of 1 leucine solute per 50 water molecules, up to a concentration of 1 leucine solute per 25 water molecules. The fact that the position of the peak does not change, yet the area under the peak increases with increasing concentration, suggests that a stable and ordered solute–solute distribution is present over a large concentration range. The experimental data over the full range of concentration, interpreted by molecular dynamics simulations that reproduce the same trends seen experimentally, appear to be inconsistent with the molecules of NALMA segregating themselves completely from the aqueous solvent to form a large hydrophobic cluster. Instead, the experiments can be interpreted as evidence for the formation of mono-dispersed to small molecular aggregates of leucine amino acids.

Experimental and computational methods

Experimental protocol

The equipment used for our X-ray scattering experiments consisted of an RU-300 Rigaku rotating anode generator fitted with a copper target, a home-built liquid sample holder with planar geometry, and a Rigaku-AXIS IV image plate detector. Yale mirrors in a helium box (Molecular Structure Corporation) and a nickel filter were used to produce a focused, monochromatic beam, which was further limited in size by a 300 μm pinhole collimator placed about 1 cm from the sample. The home-built liquid sample holder (described below) was mounted onto a standard goniometer base, allowing easy attachment to the single-axis sample goniometer. A modified lead beam stop (attached directly to the sample holder) replaced the commercial beam stop that is normally mounted on the collimator, in order to avoid small movements of the collimator during sample changes, but whose diameter prohibits data collection below Q = 0.3 Å^{-1}. The entire path length between the collimator and the sample, and between the sample and the beam stop, was purged with helium in order to minimize the continuous background due to air scattering. Apart from the liquid sample holder with its accompanying beam stop and helium enclosure, the equipment was unchanged from its standard configuration, as used for protein crystallography.

The home-built sample holder consisted of two lucite plates separated by a rubber gasket. One of the plates was fitted with two 2.5 mm stainless steel tubes, entering and leaving the sample chamber respectively, so that water or aqueous solutions could fill the sample holder after assembly. A 2 mm cylindrical hole was drilled through the back plate for the incident X-ray beam, and a cone-shaped hole, tapering at 45° to 2 mm at the sample, was machined

into the front plate. The holes were sealed with Kapton windows, 22.8 μm thick, glued to the inner face of the two lucite plates. Kapton was chosen for its rigidity, thinness, and scattering profile that can be subtracted easily from the solution scattering measurement. The sample chamber enclosed by the two plates and the rubber gasket was machined to be nominally 250 μm thick, but bowing of the Kapton windows during filling often resulted in a sample thickness closer to 300 or 400 μm.

Individual scattering patterns were recorded as 10-minute-long exposures. A measurement made with an empty sample holder was subtracted interactively from the measurement made with the filled sample holder. This was done to fully subtract the relatively sharp Bragg ring at Q = 0.402 $Å^{-1}$ produced by the Kapton window material, at the same time subtracting the residual helium and air scattering background. The liquid sample thickness was measured from the percent transmission of the primary X-ray beam, measured with a PIN diode (Molecular Structure Corporation) for the empty and for the filled sample holder. A mass absorption coefficient for water of 10.26 cm^2/g, based upon atomic values given in the International Tables of Crystallography [16], was further scaled by the measured mass density for each of our aqueous solutions. Measured values obtained at different times were corrected for variations in the incident beam intensity and in the sample thickness. All experiments included two measurements made on pure water – at the beginning of a run and again at the end. This provided an internal standard for putting the data on an absolute scale (see below) and to correct for small changes in the detector sensitivity that were observed to occur over periods of weeks or months.

The incident X-ray beam was perpendicular to the plane of the sample and to the plane of the image-plate detector, simplifying the various geometric corrections that had to be made to the data. The data were first integrated circularly within rings of constant radial increment, set to give an angular resolution of 2.62 mrad at small angle. The resulting radial scattering curves were corrected for angle-dependent absorption in the sample and in the air path between the sample and the detector. The curves were further corrected for polarization, using the function given in the International Tables of Crystallography [16] for unpolarized incident X-rays. An inverse cosine term was applied in order to correct for the flat-plate geometry of the area detector.

The relationship between scattering angle and points on the area detector was calibrated by the use of a thin copper foil inserted into the cavity of the sample holder. The observed diffraction peak (powder ring) position and the known lattice constant of copper were used to measure the sample-to-detector distance, from which the scattering angle at any point on the detector could be calculated.

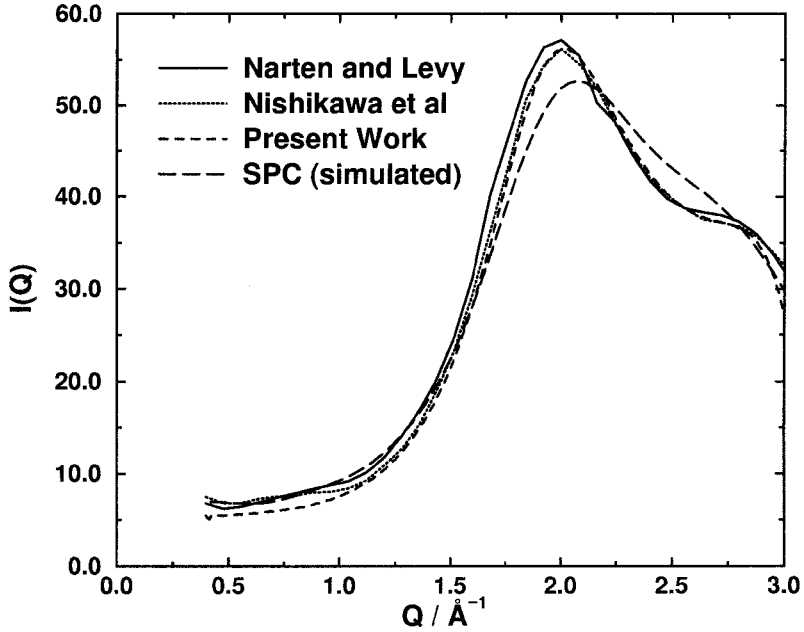

Figure 1. Experimental and simulated X-ray scattering intensity curves for pure water. The experimental curves correspond to our X-ray scattering experiment, that due to Nishikawa and Kitagawa [17], and that due to Narten and Levy [19]. The simulated scattering curve was obtained with the SPC water model [25].

Our measured scattering curves for pure water were scaled to the curve published by Nishikawa and Kitagawa [17], which are on an absolute scale, to which Compton scattering calculated according to Hajdu [18] was added in order to make the reference comparable to our experimental data. This in turn allowed us to place all other measurements on an absolute scale, since the relative scale is known between all of our measurements. After application of systematic corrections as described above, the two measurements made on pure water in the protocol of each experimental run normally agree to better than 1%. Once placed on an absolute scale, we then subtracted Compton scattering from our curves, so that comparisons can be made to scattering curves calculated from our molecular dynamics simulations of water and aqueous solutions.

As shown in Figure 1, the agreement between our measurement for pure water and the reference curve published by Nishikawa and Kitagawa [17], scaled by only a single, multiplicative factor, is quite excellent over the whole range of our measurement. There is also close agreement with the older measurement of Narten and Levy [19]. We note that the simulated SPC [25] X-ray scattering curve for pure water does not quantitatively reproduce any of the

Table 1. Simulated configurations for predicting X-ray scattering

Solute	Concentration #solutes:#waters	Preparation	N_{waters}	$N_{solutes}$
NALMA	1:24	Dispersed	362	15
	1:24	Molecular aggregates	362	15
	1:24	Hydrophobic cluster	362	15
	1:47	Dispersed	422	9
	1:47	Molecular aggregates	422	9
	1:47	Hydrophobic cluster	422	9
	1:47	Hydrophobic cluster	844	18
	1:75	Dispersed	452	6
	1:75	Backbone cluster	452	6
NALA	1:24	Dispersed	362	15
	1:47	Dispersed	422	9
	1:75	Dispersed	452	6
NAGA	1:24	Dispersed	410	17
	1:51	Dispersed	458	9
	1:79	Dispersed	476	6
NAGMA	1:25	Dispersed	400	16
	1:50	Dispersed	449	9
	1:78	Dispersed	470	6
NAQA	1:25	Dispersed	317	15
	1:48	Dispersed	431	9
	1:76	Dispersed	458	6

pure water experimental curves. In particular, the position of the main water diffraction peak for SPC is at 2.07 $Å^{-1}$, instead of the experimental value of 2.0 $Å^{-1}$, nor is there presence of a shoulder at higher Q. We return to this point later in the ‘Discussion and conclusions’ section.

Simulation protocol

Analysis of the X-ray scattering profile for each aqueous solution was aided by performing molecular dynamics simulations. The simulations were carried out at 298.15 K in the NVT ensemble using multiple time step integration of a Nosé–Hoover chain of thermostats [20,21] with a time step of 1.5 fs. Ewald

sums were used for calculation of the long-range Coulomb forces [22]. κ was set to 6.4/L, where L is the length of the simulation box, and a total of 2×292 k-vectors were used ($|k_{max}|^2 = 26$). Rigid-body dynamics for the solvent were integrated using RATTLE [23]. The AMBER force field due to Cornell et al. [24] and the SPC water model [25] were used for modeling the solutes and water, respectively.

Configurations at a given concentration were prepared by setting up the solutes in a desired configuration (see below), adding this solute configuration to a pre-equilibrated box of 512 waters (L = 24.83 Å), and deleting the appropriate number of overlapping waters for the concentration and density being simulated. Table 1 shows the resulting numbers of solute and water molecules in each simulation box. These configurations were relaxed by low-temperature quenching, and equilibrated for 75 ps before any statistics were collected.

For most of the studies the initial solute configuration was constructed to be as dispersed as possible. This was set up by changing all charges on the solutes to be of the same sign and running the resulting simulation at low temperature in the gas phase (with no water present). The net effect of this procedure is to set up the solutes in a maximally dispersed configuration for that concentration. The normal AMBER parameters were there turned back on, the configurations were relaxed by low-temperature quenching to allow the solutes to regain their shape, and the solute configurations were placed in aqueous solution as described above. These simulations were run for 75 ps to collect statistics after the initial 75 ps equilibration period.

For the analysis of the small-angle feature in the NALMA scattering curve, we performed further simulations of various representative configurations: dispersed, small molecular aggregates, and fully clustered (Table 1). The degree to which the solute configurations were dispersed, aggregated, or fully clustered can be quantified using a solute centers distribution function generated during the simulations. Figure 2 shows the centers radial distribution function, $g_c(r)$, at a concentration mole ratio of solute to water of 1:24 for NALMA molecules when they are configured as dispersed, forming small aggregates, and fully clustered. The dispersed $g_c(r)$ has no well-developed peak at any relative separation, while the fully clustered configuration has three well-defined peaks consistent with a dense liquid of NALMAs. The molecular aggregates $g_c(r)$ is clearly more structured than the dispersed configuration, less structured than the full cluster and instead is consistent with the formation of smaller, loosely associated clusters.

Maximally dispersed configurations were prepared as described above, but the time periods for equilibration and statistics collection were shortened (30 ps for both) to allow the configurations to maintain their dispersed quality.

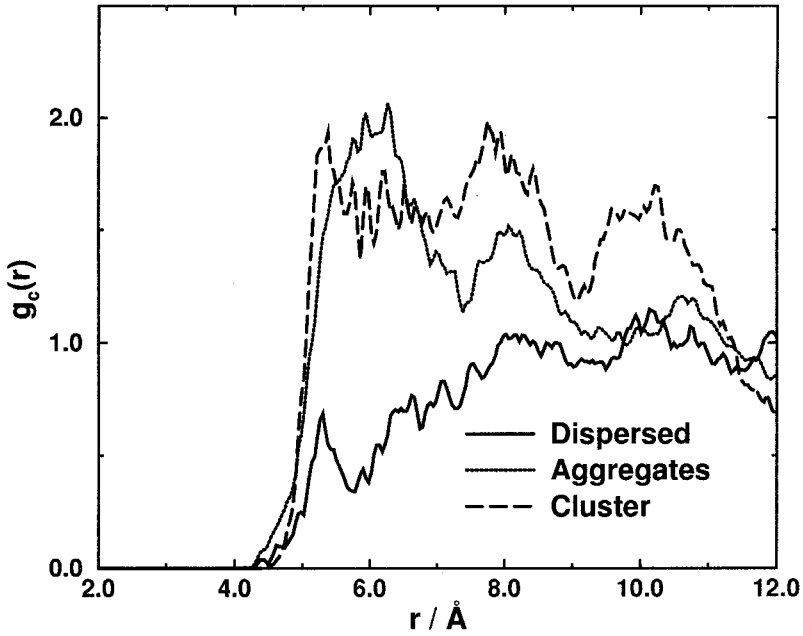

Figure 2. The solute centers radial distribution function, $g_C(r)$, at a mole ratio of solute to water of 1:24 for NALMA molecules when configured as dispersed, forming small aggregates, and fully clustered.

Three such configurations were prepared and simulated; results from these three simulations were averaged together to give the dispersed scattering predictions. The configuration of molecular aggregates for 1:24 NALMA was prepared by running a 450 ps simulation at this concentration. At the end of this run, the solutes had grouped into clusters of two to six molecules each, and this configuration was run for a further 75 ps to gather statistics. Cluster configurations were prepared by enhancing the Lennard-Jones ϵ value on the γ-carbon of the leucine side chains and running low-temperature gas-phase simulations of the solutes alone. The resulting solute configurations represented clusters driven together by hydrophobic association and were placed in aqueous solution and simulated as described above.

Predicted scattering

X-ray scattering can be predicted from the atom–atom radial distribution functions available in the simulations. The total scattering is a function of solute–solute, solute–water, water–water, and intramolecular correlations:

$$I_{solution}(Q) = I_{solute-solute}(Q) + I_{solute-water}(Q) + I_{water-water}(Q) + I_{intra}(Q) \quad (1)$$

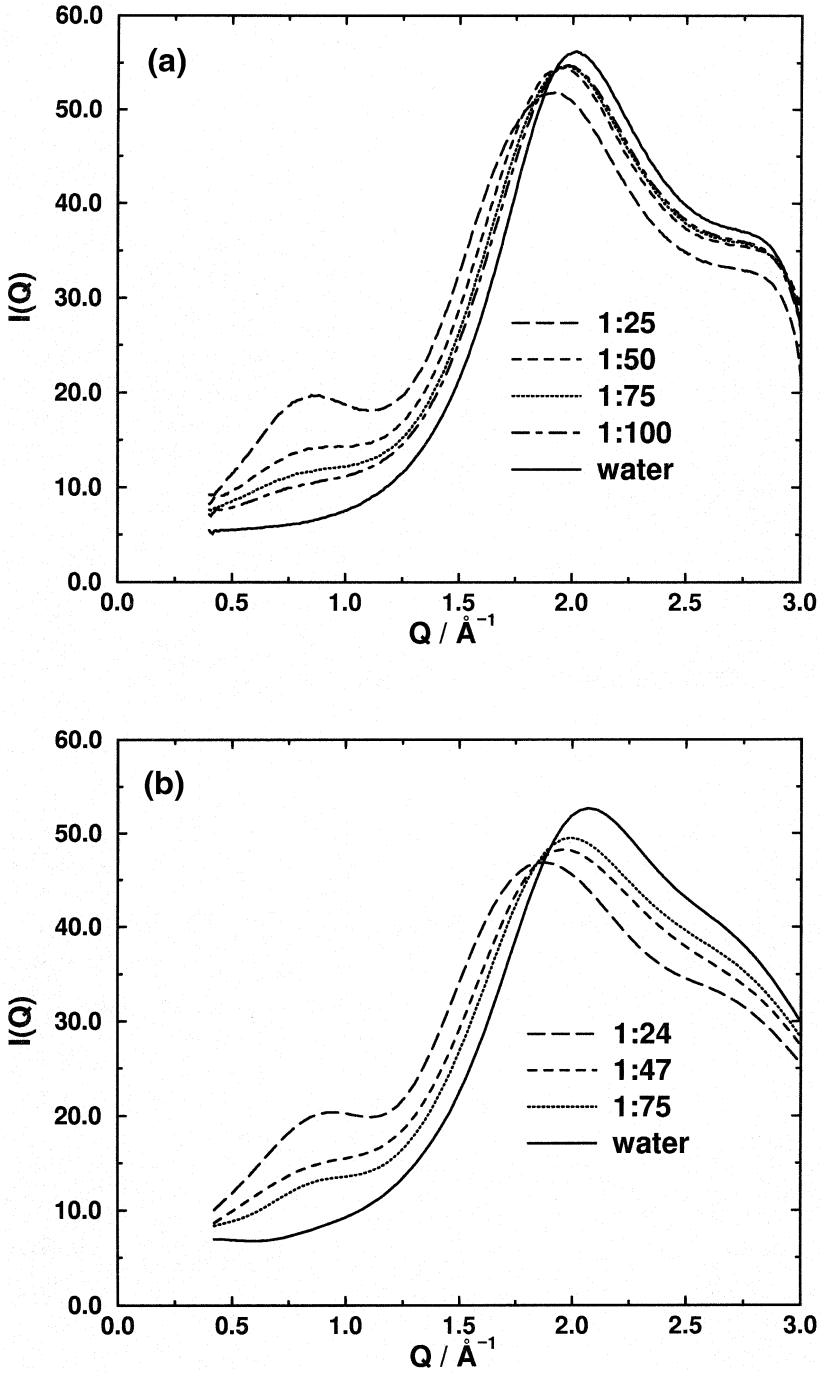

Figure 3. Experimental and simulated X-ray scattering intensity curves for pure water and NALMA in water solutions (mole ratios of solute to water of ∼1:25, ∼1:50, ∼1:75, and 1:100 for experiment).

For predicting the scattering from a heterogeneous group of molecules, as in the present case, it is easiest to view a given volume of aqueous solution as a collection of atoms, and calculate the resulting scattering from this point of view [26,27]:

$$I(Q) = \sum_i x_i f_i(Q)^2 + \rho \sum_{i \leq j} x_i x_j f_i(Q) f_j(Q) \left[\hat{h}_{ij}(Q) + \hat{s}_{ij}(Q)\right](2 - \delta_{ij}) \qquad (2)$$

where the sums are over the atom types present; x_i is the mole fraction of atom i; $f_i(Q)$ is the atomic structure factor [16] for atom i; ρ is the atomic density; $\hat{h}_{ij}(Q)$ is the Fourier transform of $h_{ij}(r) = g_{ij}(r) - 1$; $\hat{s}_{ij}(Q)$ is the intramolecular correlation function for atoms i and j in Q-space [28]; and δ_{ij} properly counts each term for the sum over $i \leq j$. The resulting scattering predictions are good only for $Q \geq \sim 0.25$ Å^{-1} since the finite size of the simulation box prevents us from studying longer-ranged correlations. The solute intramolecular contribution to scattering was calculated differently from our previous work [10–13]; we used the solute conformations from simulation to derive this term instead of using averaged crystal structures. This was possible in the present work because of the extra statistics provided by having many solutes present in the simulation box. The results for the predicted solute intramolecular contributions were not significantly different between the two approaches.

Results

Figures 3–6 display the X-ray solution scattering measurements (Figures 3a–6a) and molecular dynamics simulations (Figures 3b–6b) for the blocked amino acids NALMA, NALA, NAQA, and NAGMA, over a range of concentrations of solute to water mole ratios of 1:25 to 1:100, as well as for pure water. We note that NALA is not soluble for concentrations greater than 1:35, and, for reasons discussed later, we also simulated the X-ray scattering for NALA at a mole ratio of solute to water equal to 1:16 (Figure 4b). The choice of these particular amino acids is motivated by two reasons: (i) they are soluble in water over a sufficient range of concentrations; and (ii) for the purposes of understanding differences in hydration properties, they are representative hydrophobic and hydrophilic side chains, and model peptide backbone, respectively. We also evaluate the differences in observed scattering that depends on the nature of the blocking group, by considering both NALMA and NALA.

Figure 3a shows the experimental results for NALMA at ratios of solute to water equal to 1:100, 1:75, 1:50, and 1:25, as well as the results for pure

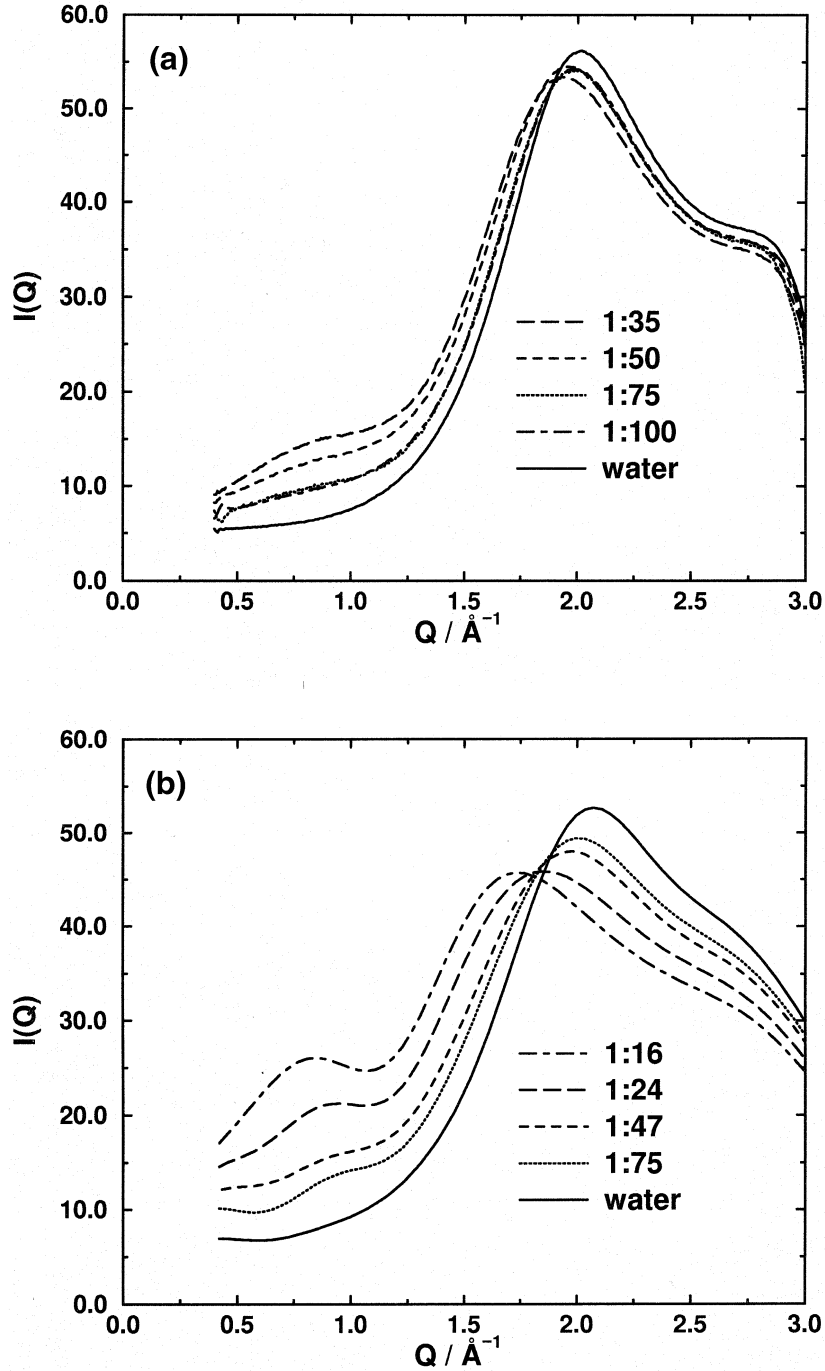

Figure 4. Experimental and simulated X-ray scattering intensity curves for pure water and NALA in water solutions (mole ratios of solute to water ranging from ∼1:16 to ∼1:75, and 1:100 for experiment).

water. The X-ray curve for the more dilute concentrations is dominated by the main X-ray diffraction peak of water which occurs at Q $\sim$ 2.0 Å^{-1} at room temperature, although that peak is modified with respect to pure water by being shifted to slightly smaller angle and is of lesser intensity. However, at a concentration of 1:50, the main diffraction peak continues to shift to smaller angle, while a new feature appears at Q $\sim$ 0.8 Å^{-1}. At the highest concentration of 1:25, the main water peak loses substantial intensity and shows a further shift to smaller Q-values, while the feature at Q $\sim$ 0.8 Å^{-1} develops into a clearly defined maximum.

Figure 4a shows the experimental results for NALA at ratios of solute to water equal to 1:100, 1:75, 1:50, and 1:35, as well as for pure water. The only difference in the two leucine solutes studied here is the substitution of a hydrogen for the methyl group on the terminating backbone amide. By comparing NALMA and NALA, we see that the shift and decrease in intensity of the main diffraction peak, and the feature at Q $\sim$ 0.8 Å^{-1} at more concentrated solutions, is present but diminished with respect to the NALMA results. The diminished effect is due in part because NALA is not soluble at 1:25, and comparisons cannot be made to the NALMA solution at the concentration in which the effect is greatest. However, the hydrophobic methylamide group clearly plays a role in the hydration structure, and this point is considered further below.

Figure 5a displays the experimental results for NAQA at the same mole ratios of solute to water as for NALMA, as well as for pure water. Again, the X-ray curve for the more dilute concentrations is dominated by the main X-ray diffraction peak of water at room temperature at Q $\sim$ 2.0 Å^{-1}, although intensity is lost with respect to pure water. While the peak position of the main water peak shifts to smaller angle, it does not, however, lose further intensity over the full concentration range. More importantly, a peak at Q $\sim$ 0.8 Å^{-1} does not occur even at the highest concentration of 1:25, although there is a general rise in the small-angle baseline. Even though NAQA is of similar size to NALA, it seems apparent that the hydrophilic character of the glutamine side chain contributes to a different distribution of solute configurations than is generated by the hydrophobic amino acid.

Figure 6a displays the experimental results for NAGMA at the same mole ratios of solute to water as NALMA and NAQA, as well as for pure water. The main X-ray diffraction peak of water at room temperature at Q $\sim$ 2.0 Å^{-1} again shifts to smaller angle, but much less so than for the larger amino acids. Furthermore, the main water peak does not lose significant intensity with respect to pure water, and there is no appearance of a peak at Q $\sim$ 0.8 Å^{-1}, although there is a small rise in the small-angle baseline at the highest con-

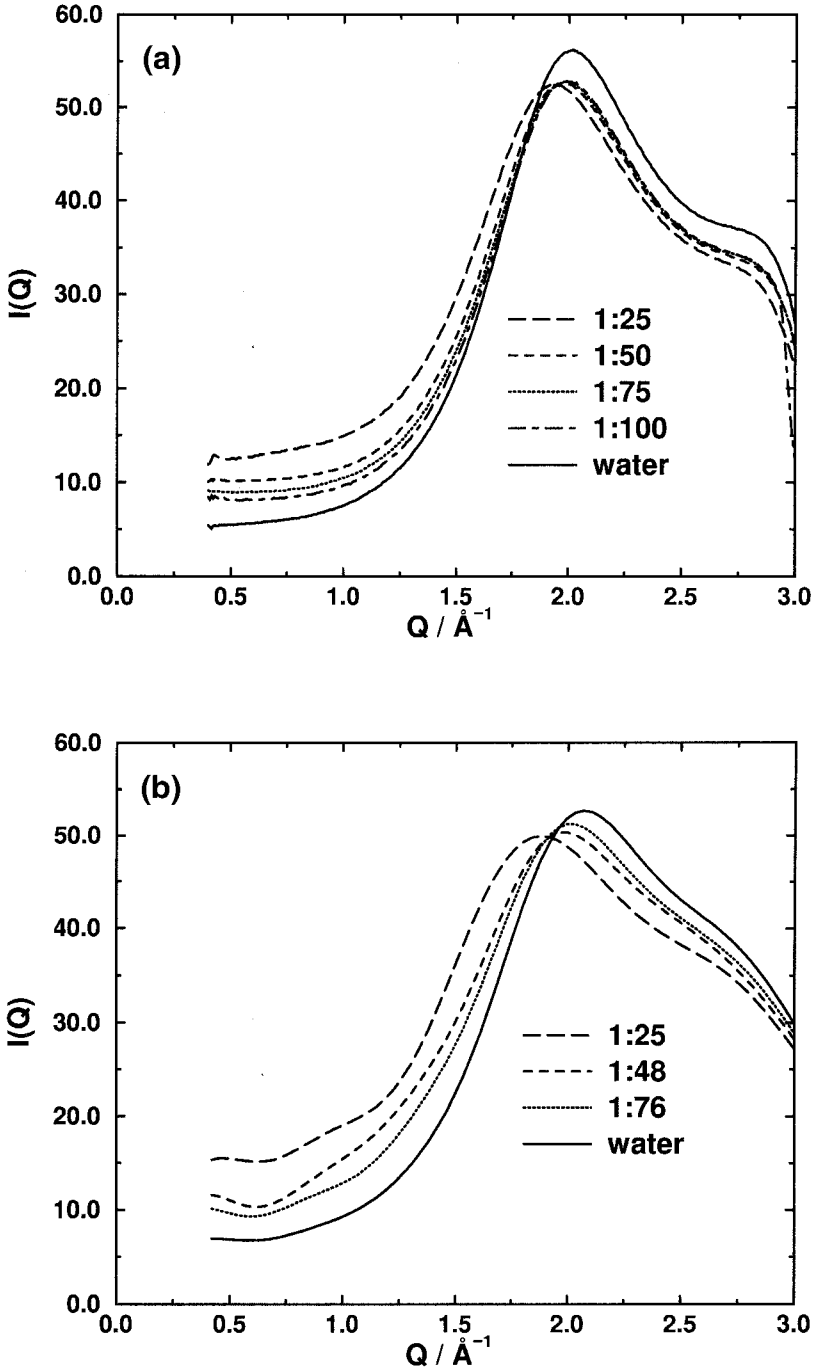

Figure 5. Experimental and simulated X-ray scattering intensity curves for pure water and NAQA in water solutions (mole ratios of solute to water of ∼1:25, ∼1:50, ∼1:75, and 1:100 for experiment).

centration of 1:25. The NAGMA results help to distinguish side chain versus backbone effects of solvation.

It is quite clear that the more concentrated solutions of NALMA exhibit a novel peak at Q $\sim$ 0.8 Å^{-1}, reflecting the formation of a fluid and partially ordered phase (Figure 3). The new diffraction peak at Q $\sim$ 0.8 Å^{-1} shifts negligibly at the two measured concentrations of 1:50 and 1:25, which indicates that the effective length scale represents a stable solute–solute configuration. Although the internal length scale in this configuration is not sensitive to solute concentration, unlike the peak observed for concentrated solutions of $NiCl_2$ [29], the amount of material assembled at that length scale, as judged by the area under the peak, depends steeply on concentration. When the results for NALMA are compared to those for NALA, we also see that the methyl blocking group contributes to an increase in intensity at Q $\sim$ 0.8 Å^{-1}.

We use molecular dynamics simulations to interpret these trends. The new experimental feature at Q $\sim$ 0.8 Å^{-1} is replicated by simulation for NALMA (Figure 3b) and a comparison is also made to simulated X-ray scattering curves for NALA (Figure 4b), NAQA (Figure 5b), and NAGMA (Figure 6b). It is important to emphasize that we are unlikely to simulate the time progression involved in the formation of solute distributions seen experimentally, as this would require molecular dynamics simulations over very long time scales, and/or other ensembles, in order to reach the final equilibrated distribution of solutes. However, considerations of the mechanisms of how these solute configurations are reached are not important for this experiment. What is important is to determine the types of final configurations of solutes that reproduce the static experimental observable.

Table 1 lists the simulations performed to analyze the experimentally observed X-ray scattering. For each of the blocked amino acids NALMA, NALA, NAQA, and NAGMA, we have considered a fully dispersed and hydrated configuration of solute molecules at mole ratios of solute to water of $\sim$1:24, $\sim$1:50, and $\sim$1:75. For NALMA and NALA at the more concentrated solutions (1:24 and 1:16), solute distributions were simulated in which mono-dispersed to small molecular aggregates are observed. Additionally for NALMA, fully clustered configurations in which the solutes formed a cylinder with a well-packed hydrophobic core, and a larger cluster that was spherical in shape with hydrophobic side chain interactions dominating the core and exposing hydrophilic groups at the surface, were simulated. The $g_c(r)$ distribution functions that were generated from the simulations (listed in Table 1) showed that the solutes indeed maintained a desired class of solute distribution: maximally dispersed, single solutes and small molecular aggregates, and fully clustered.

It is clear from Figures 3b–6b, when compared against their corresponding experimental partner (Figures 3a–6a), that the trends with respect to amino

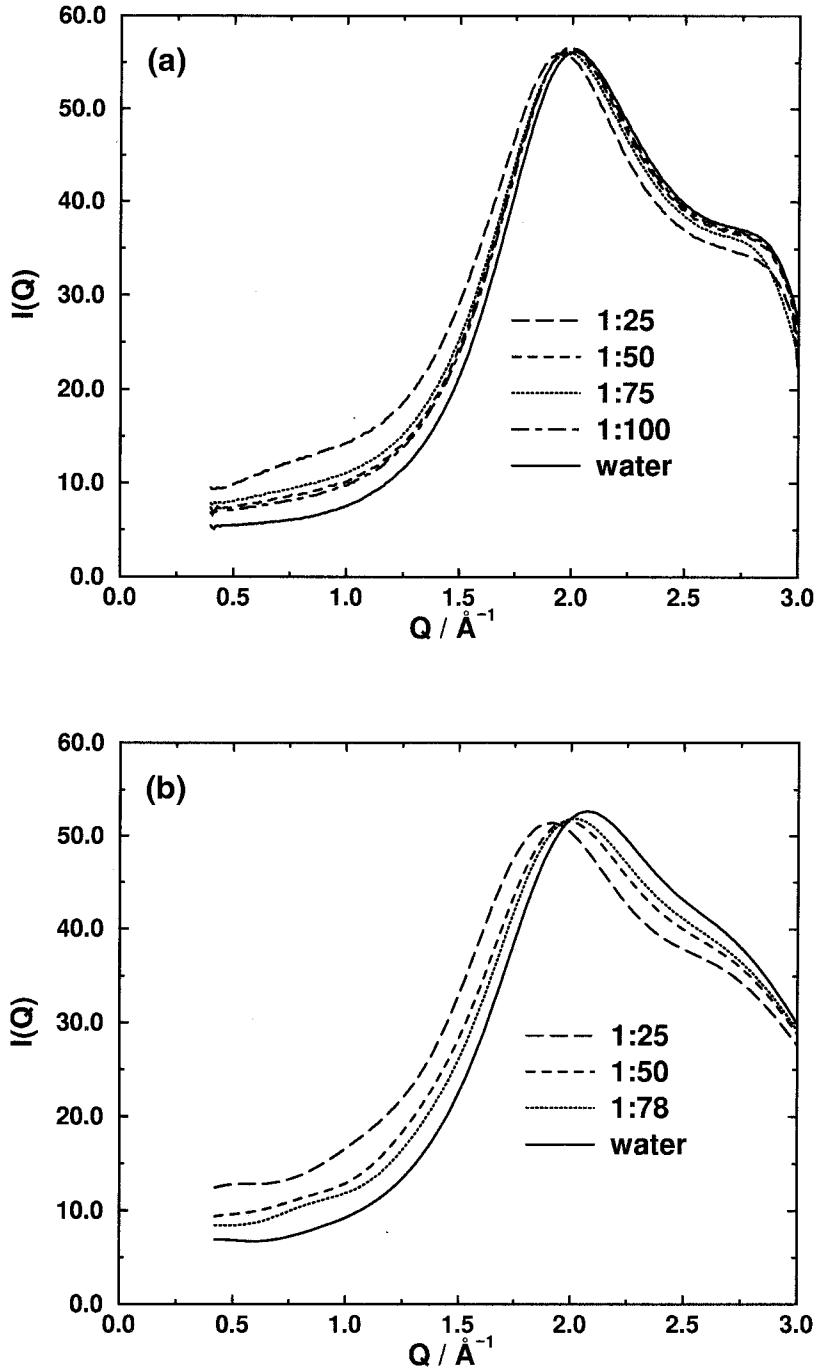

Figure 6. Experimental and simulated X-ray scattering intensity curves for pure water and NAGMA in water solutions (mole ratios of solute to water of ∼1:25, ∼1:50, ∼1:75, and 1:100 for experiment).

acid type and/or blocking group are well reproduced by the simulations of the fully dispersed or molecular aggregate configurations for the more concentrated solutions. In all cases, the simulations reproduce the trends in the X-ray curve as a function of concentration, although the trends are more exaggerated in the simulations. Dilute solutions show a main X-ray diffraction peak of water that is shifted to smaller angle and is of lesser intensity with respect to pure water, and more concentrated solutions show the differences between leucine, glutamine, and glycine at Q $\sim$ 0.8 Å^{-1}, in which only NALMA shows a well-defined maximum, in agreement with experiment.

Next we consider the sensitivity of the new feature for NALMA at Q $\sim$ 0.8 Å^{-1} to solute configurations. Figures 7a–c show the intensity curves derived from the simulations of the fully dispersed, small molecular aggregates, and single cluster simulations of NALMA solutes in water, respectively, as a function of concentration. We find that the simulated intensity curves do not distinguish between the maximally dispersed and small molecular aggregate configurations, but both are well differentiated from the intensity arising from a single NALMA cluster in water. Furthermore, the experimental data is best reproduced by simulated configurations of NALMAs in which the solutes are maximally dispersed or are involved in making small molecular aggregates. When considering the single cluster data (Figure 7c), the scattering predicted for the smallest single cluster is too sharply defined and slightly shifted to a smaller Q-value. This gets worse for the larger-sized cluster (which is simulated in a larger box and is therefore more dilute) where there is a significant shift to smaller Q.

Given the rather good agreement in trends between simulation and experiment, we can extrapolate beyond the experimental solubility range of NALA, to concentrations of both 1:24 and 1:16 solute to water ratios by simulation (Figure 4). We believe that NALA is less soluble in water than NALMA because of the stronger hydrogen-bonding interactions of the amide group for NALA in the crystal, which is weakened with the methylamide blocking group of NALMA. We see that NALA also shows the same evolution from shoulder to peak, but over a different concentration range than NALMA, requiring a concentration mole ratio of NALA solute to water of 1:24 before the peak is fully developed at Q $\sim$ 0.8 Å^{-1}. This indicates that the underlying effect does not require the methylamide blocking group, but it is suggestive that the degree of hydrophobic aggregation which is involved in the creation of the new effective Bragg spacing is ‘accumulative’ over the number of hydrophobic entities present in the solute molecule. This point is accentuated by the comparison of the NALMA and NALA results (Figures 3 and 4), which shows that the methyl blocking group does contribute to a sharpening of the main water diffraction peak, as well as the greater intensity at Q $\sim$ 0.8 Å^{-1},

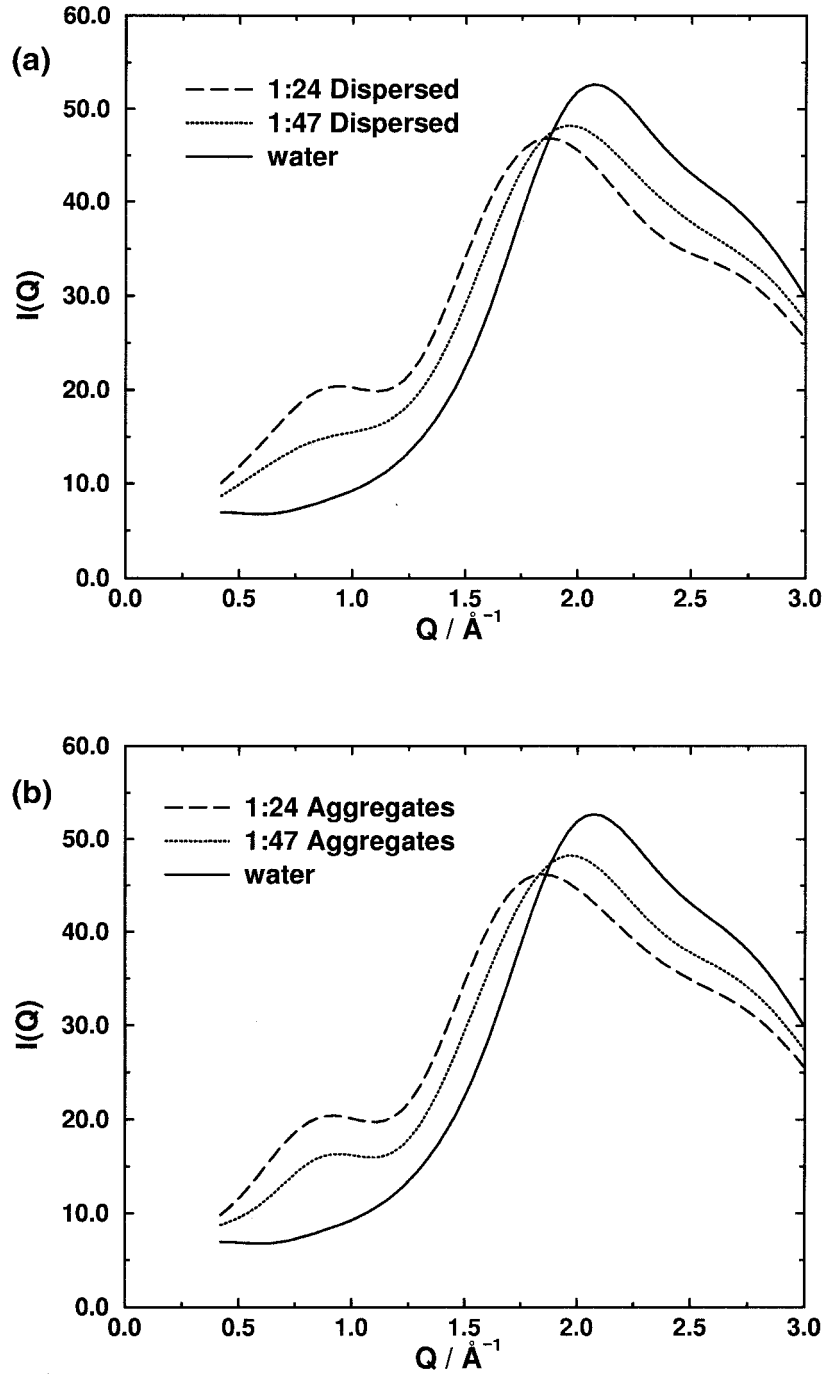

Figure 7. Simulated X-ray scattering intensity curves for pure water and NALMA in water (a) at mole ratios of solute to water of 1:24 and 1:47, with NALMAs maximally dispersed; (b) at mole ratios of 1:24 and 1:47, with NALMAs configured as a distribution of small molecular aggregates.

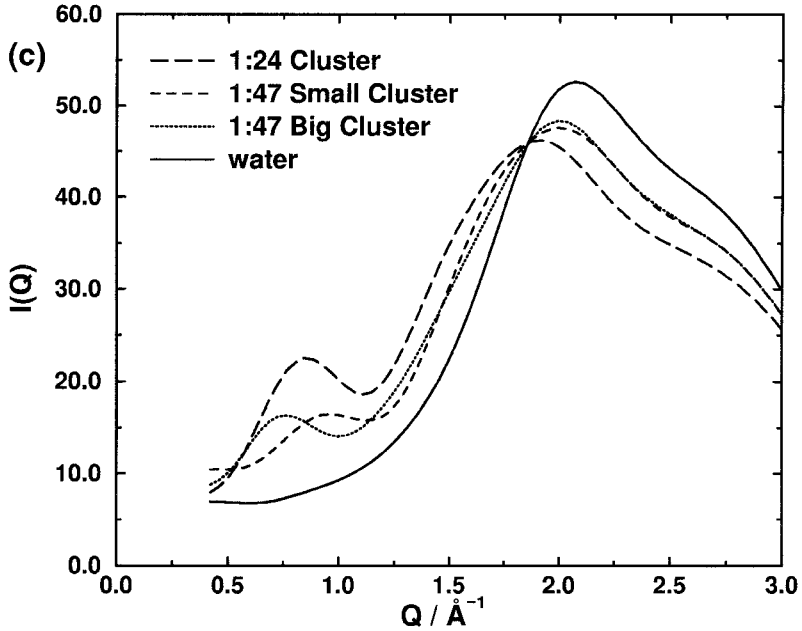

Figure 7. (c) Simulated X-ray scattering intensity curves for pure water and NALMA in water configured as a single cluster, at mole ratios of 1:24, 1:47 (small cluster) and 1:47 (big cluster).

at higher concentrations. A more detailed structural analysis of the series of experiments reported here will be provided in a future publication [30].

Discussion and conclusions

We have designed X-ray solution scattering and molecular dynamics simulation studies to characterize hydration as a function of concentration for the amino acids leucine, glutamine, and glycine in either *N*-acetyl-amino acid-amide or -methylamide forms. As the concentration increases, we observe a new diffraction peak that grows in at Q $\sim$ 0.8 Å^{-1} for only the hydrophobic amino acid leucine. The formation of a distinct diffraction peak with increasing solute concentration suggests the formation of a stable and partially ordered leucine solute–solute distribution, something that is not observed for either glutamine or glycine. The experimental results were analyzed by simulating various distributions of solutes in water, ranging from maximally dispersed to fully clustered solutes. We find that the experimental results for glutamine and glycine are best reproduced by simulations during which these amino acids are maximally dispersed. For concentrated solutions of the hydrophobic leucine solutes NALMA and NALA, our experimental and

simulated solution scattering results support a view that small hydrophobic domains are observed and therefore sustained in preference to single, large clusters that expose a hydrophilic surface and segregated hydrophobic core.

We view our solution scattering experiments and simulations as a model systems approach, similar to the determination of isolated secondary structure elements that might serve as folding intermediates [31–37], but in the realm of hydration forces for solutes with full amino acid complexity [13]. The solution scattering experiments and simulations presented here suggest a model of later protein folding events when significant spatial domains of the protein comprise a hydrophobic core. Various interactions [38–43] such as van der Waals and electrostatics arising from the hydrophilic and hydrophobic character of the polypeptide chain, and differences in amino acid side chain entropy costs, are all important at these molecular length scales, a complexity which is often ignored with an assumption that purely hydrophobic effects dominate.

Calculations of the affinities with which drug molecules bind to a given gene product must also account for the influence that the aqueous solvent may have on the mediation of drug–target interactions. The geometry and thermodynamics of water binding is difficult to model because water is an associated liquid whose hydrogen-bonded network of molecules, while adaptive to perturbations in temperature, pressure, or introduction of a foreign solute, does so with unusual structural and/or thermodynamics consequences [44,45]. However, the versatility of intermolecular interactions between ligand and target, and therefore conclusions about a particular drug's binding affinity, can be completely altered with the inclusion of water into the binding affinity estimate [1].

Empirical protein force fields are often used in conjunction with free energy perturbation techniques, or as a direct scoring function for a docked structure, to ascertain the relative binding energy for either changes in docking geometry or among drug candidates. X-ray (and neutron) solution scattering measurements, such as those reported here, can provide an independent check on the accuracy of simulation parameters that are used for aqueous hydration free energies and the structural properties of amino acids in water. In particular, we find that the simulated SPC X-ray scattering intensity for pure water does not adequately reproduce our experimental data curve. Preliminary analysis seems to suggest that SPC is 'under-structured', in that the position of the main water diffraction peak is at 2.07 $Å^{-1}$, instead of the experimental value of 2.01 $Å^{-1}$. We find that other water models that have a more structured $g_{OO}(r)$, in particular, better reproduce our experimental scattering on pure water [46]. Similarly, our simulated solution scattering measurements tend to be 'over-structured', in that the solute induces too large a shift of

the main water diffraction peak to smaller angle, possibly overemphasizing what is at best weak dewetting (expansion of the effective Bragg spacing of water) when hydrophobic groups are present. Further structural analysis of the simulations and experiment must be considered before such conclusions can be firmly made [30]. Ultimately, existing protein and water force fields can be tested and modified when necessary, until the simulated neutron and X-ray scattering profiles quantitatively reproduce experiment for a large variety of biologically relevant solutes in water.

In turn, a better reproduction of the experimental scattering curve from simulation, using improved force fields, can ultimately lead to the extraction of implicit hydration potentials of mean force from the scattering experiment. In recent work, we have subtracted from a neutron solution scattering signal simulated quantities that describe uncorrelated NALA solutes in water, to leave an excess signal that contains information about the correlated solutes in water at dilute concentration [12]. Various model pair distribution functions for NALA molecules (gas, cluster, and aqueous forms of $g_c(r)$) were tested for their ability to reproduce this excess experimental signal. We found that the excess experimental signal is adequate enough to rule out gas and cluster pair correlation functions, and that the aqueous form adequately reproduced the excess experimental signal [12]. Improvements in explicit solute and water force fields would allow us to isolate a more robust experimental signal arising from the solute centers pair distribution function, $g_c(r)$. The importance of $g_c(r)$, which in turn defines a potential of mean force, $W(r)$, is that it describes the net correlations between solute pairs that implicitly account for the solvent environment. One possible outcome is that the derived potentials of mean force, or 'implicit' hydration potentials, could be interfaced with empirical protein force fields to be used in computational studies of protein–ligand docking and scoring.

Acknowledgements

J.M.S. is supported by a National Science Foundation Graduate Research fellowship. J.M.S. and T.H.G. thank the LDRD program through NERSC, U.S. Department of Energy contract number DE-AC-03-76SF00098, for support in FY98. G.H. and T.H.G. thank the Office of Biological and Environmental Research (OBER), U.S. Department of Energy Contract Number DE-AC-03-76SF0098 for support in FY99. T.H.G. gratefully acknowledges support from the Air Force Office of Sponsored Research, Grant No. FQ8671-9601129, and the National Energy Research Supercomputer Center for computer time. We thank Dr. Bing Jap for use of his X-ray machine to obtain our X-ray solu-

tion scattering data, Rick Burkhard for help with the sample holder design, and Alex Pertsemlidis for use of his circular integration code.

References

1. Ladbury, J.E., Chem. Biol., 3 (1996) 973.
2. Laskowski, R.A., Thornton, J.M., Humblet, C. and Singh, J., J. Mol. Biol., 259 (1996) 175.
3. Klebe, G., J. Mol. Biol., 237 (1994) 212.
4. Hobohm, U., Scharf, M., Schneider, R. and Sander, C., Protein Sci., 1 (1992) 409.
5. Eisenberg, D. and McLachlan, A.D., Nature, 319 (1986) 199.
6. Makhatadze, G.I. and Privalov, P.L., J. Mol. Biol., 232 (1993) 639.
7. Murphy, K.P. and Gill, S.J., J. Mol. Biol., 222 (1991) 699.
8. Honig, B. and Nicholls, A., Science, 268 (1995) 1144.
9. Head-Gordon, T., Proc. Natl. Acad. Sci. USA, 92 (1995) 8308.
10. Pertsemlidis, A., Saxena, A.M., Soper, A.K., Head-Gordon, T. and Glaeser, R.M., Proc. Natl. Acad. Sci. USA, 93 (1996) 10769.
11. Head-Gordon, T., Sorenson, J.M., Pertsemlidis, A. and Glaeser, R.M., Biophys. J., 73 (1997) 2106.
12. Pertsemlidis, A., Soper, A.K., Sorenson, J.M. and Head-Gordon, T., Proc. Natl. Acad. Sci. USA, 96 (1999) 481.
13. Sorenson, J.M., Hura, G., Soper, A.K., Pertsemlidis, A. and Head-Gordon, T., J. Phys. Chem. B, 103 (1999) 5413.
14. Pertsemlidis, A., Ph.D. Thesis, University of California at Berkeley, 1995.
15. Sorenson, J.M. and Head-Gordon, T., Fold Design, 3 (1998) 523.
16. Ibers, J.A. and Hamilton, W.C., International Tables for X-ray Crystallography, Vols. 2–4, 3rd edition, Kluwer Academic Publishers, Dordrecht, 1989.
17. Nishikawa, K. and Kitagawa, N., Bull. Chem. Soc. Jpn., 53 (1980) 2804.
18. Hadju, F., Acta Crystallogr., A28 (1972) 250.
19. Narten, A.H. and Levy, H.A., J. Chem. Phys., 55 (1971) 2263.
20. Martyna, G.J., Tuckerman, M.E., Tobias, D.J. and Klein, M.L., Mol. Phys., 87 (1996) 1117.
21. Nosé, S., In Meyer, M. and Pontikis, V. (Eds) Computer Simulations in Material Science, Kluwer Academic Publishers, Dordrecht, 1991, p. 21.
22. Allen, M.P. and Tildesley, D.J., Computer Simulation of Liquids, Clarendon Press, Oxford, 1987.
23. Andersen, H.C., J. Comput. Phys., 52 (1983) 24.
24. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A., J. Am. Chem. Soc., 117 (1995) 5179.
25. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermanns, J., In Pullman, B. (Ed.) Intermolecular Forces, Reidel, Dordrecht, 1981, p. 331.
26. Magini, M., Licheri, G., Paschina, G., Piccaluga, G. and Pinna, G., X-ray Diffraction of Ions in Aqueous Solution: Hydration and Complex Formation, CRC Press, Boca Raton, FL, 1988.
27. Egelstaff, P.A., An Introduction to the Liquid State, Oxford Series on Neutron Scattering in Condensed Matter, Oxford University Press, Oxford, 1992.

28. Hansen, J.P. and McDonald, I.R., Theory of Simple Liquids, 2nd edition, Academic Press, London, 1986.
29. Powell, D.H., Neilson, G.W. and Enderby, J.E., J. Phys. Condens. Matter, 1 (1989) 8721.
30. Sorenson, J.M., Hura, G. and Head-Gordon, T., in preparation.
31. Wright, P.E., Dyson, H.J. and Lerner, R.A., Biochemistry, 27 (1988) 7167.
32. Dyson, H.J., Rance, M., Houghten, R.A., Lerner, R.A. and Wright, P.E., J. Mol. Biol., 201 (1988) 161.
33. Marqusee, S. and Baldwin, R.L., In Gierasch, L.M. and King, J. (Eds.) Protein Folding: Deciphering the Second Half of the Genetic Code, American Association for the Advancement of Science, Washington, DC, 1990, p. 85.
34. Wright, P.E., Dyson, H.J., Waltho, J.P. and Lerner, R.A., In Gierasch, L.M. and King, J. (Eds.) Protein Folding: Deciphering the Second Half of the Genetic Code, American Association for the Advancement of Science, Washington, DC, 1990, p. 85.
35. Tobias, D.J., Sneddon, S.F. and Brooks III, C.L., J. Mol. Biol., 216 (1990) 783.
36. Tobias, D.J. and Brooks III, C.L., Biochemistry, 30 (1991) 6059.
37. Tobias, D.J., Sneddon, S.F. and Brooks III, C.L., J. Mol. Biol., 227 (1992) 1244.
38. Hummer, G., Garde, S., Garcia, A.E., Paulaitis, M.E. et al., J. Phys. Chem., B102 (1998) 10469.
39. Hummer, G. and Garde, S., Phys. Rev. Lett., 80 (1998) 4193.
40. Lum, K., Chandler, D. and Weeks, J.D., submitted.
41. Rank, J.A. and Baker, D., Protein Sci., 6 (1997) 347.
42. Ben-Naim, A., J. Phys. Chem., 94 (1990) 6893.
43. Dill, K.A., Biochemistry, 29 (1990) 7133.
44. Franks, F., Water, A Comprehensive Treatise, Plenum, New York, NY.
45. Stillinger, F.H., Science, 209 (1980) 451.
46. Rick, S.W., Stuart, S.J. and Berne, B.J., J. Chem. Phys., 101 (1994) 6141.